Working Paper No. 136

AD 670734

EFFICIENT CHOICE OF INFORMATION SERVICES

by

Jacob Marschak

May 1968

Operations Research Division

WESTERN MANAGEMENT SCIENCE INSTITUTE
University of California, Los Angeles

30

University of California

Los Angeles

Western Management Science Institute

Working Paper No. 136

"EFFICIENT CHOICE OF INFORMATION SERVICES"

by

Jacob Marschak

May, 1968

Prepared for the Conference for Research on Management Information Systems Carnegie-Mellon University, Pittsburgh, 17-19 June, 1968. Supported by a grant from the Ford Foundation to the Graduate School of Business Administration, University of California at Berkeley, and administered through the Center for Research in Management Science.

ABSTRACT

An <u>information system</u> is defined as a chain of <u>information services</u>: <u>inquiring</u> -- <u>data-storing</u> -- <u>encoding</u> -- <u>transmitting</u> -- <u>decoding</u> -- <u>deciding</u>. Each is a transformer represented, in general, by a stochastic matrix and a cost function. The inputs of "inquiring" are the benefit-relevant <u>events</u> (possibly statistical parameters). <u>Actions</u> are outputs of "deciding." Together, actions and events determine the <u>benefits</u>. Other outputs of a service are: (a) inputs into the successive service, and (b) contributions to the <u>cost</u> of acquiring and operating the information system.

The decision theory of economists and statisticians has usually neglected the subsequence "data-storing -- encoding -- transmitting -- decoding." Communication engineers, on the other hand, have neglected the inquiring and deciding services and have usually equated benefit with the non-occurrence of error in the communication of data. With data pre-stored, long sequences of messages can be communicated without prohibitive delays; and useful asymptotic properties of the "information amount transmitted" and the "channel capacity" follow. These quantities are relevant to the communication cost but neither to the cost nor the benefit of inquiring and deciding.

Suppose the utility to the "manager" (the "organizer," the "meta-decider") is known to be additive in benefit and cost (both appropriately scaled), and his "prior" probability of events is known. Then, and only then, the ("efficient") subset of all feasible information systems for which the pair "expected benefit, expected cost" is not dominated by that of any other system, will contain all optimal systems. An optimal system can then be determined by a manager compelled to search for, and to apply, his "scaling functions" expressing benefits and costs in the same units.

Correspondingly, pure communication theory has assumed, in effect, utility to be additive in the following criteria (all undesirable, costly, or delay-producing): occurrence of communication error; length of code word; size of

code; and channel capacity. However, for the efficient choice of the total chain of information services, factors determining the cost of inquiring (e.g., sample size) and of deciding (e.g., computer memory) must also be considered, each properly transformed to become an additive component of utility; and an (additive) overall benefit must replace the criterion of "communication error."

CONTENTS

EFFICIENT CHOICE OF INFORMATION SERVICES

Jacob Marschak

Western Management Science Institute, University of California, Los Angeles,
and Center for Research in Management Science, University of California, Berkeley

1. INTRODUCTION

1.1. This is an attempt to clear up important misunderstandings and to achieve conceptual unity between, on the one hand, the economists and statisticians concerned with efficient decision and organization and, on the other, the communication engineers who have created what has come to be called information theory. Related thought of workers in the logic and psychology of language and of problem-solving should also find its place in the common conceptual framework.

1.2. The manager buys instruments or hires services. The distinction is not relevant for the general statement of our problem. We shall therefore speak, for brevity, of services only, with the understanding that in any particular application the size of a stock of instruments will be carefully distinguished from the number of machine-hours (or man-hours) of a service.

1.3. The term <u>information system</u> will denote the sequence of <u>information services</u>, viz., the services of <u>inquiring</u>, <u>communicating</u>, and <u>deciding</u>, in that order. More precisely, communication is itself a sequence of <u>encoding</u>, <u>transmitting</u>, and <u>deciding</u>. There is also another component of the sequence, called <u>storing</u>. It can be intermediate between any two consecutive information services and in particular between inquiring and encoding.

Each information service can be regarded as a transformer of its characteristic inputs into outputs. On Figures 1,, 2., 3., transformers are boxes; variables (sets of values of inputs or outputs) are circles. Variables are denoted by lower case Latin letters. Transformers (functions) are denoted by Greek letters, with the exception of encoding and decoding.

2. INQUIRING AND DECIDING, AT CONSTANT COST

2.1. Figure 2, "Inquiring, Communicating, Deciding," is more complete than either Figure 1, "Inquiring and Deciding," or Figure 3, "Communication only." But it will be convenient to start with Figure 1, which omits the communication aspect, and considers only two information services, "inquiring" and "deciding." We shall also disregard, for a moment, the symbols k, κ_λ, κ_α, all referring to cost. We consider the information system consisting of two consecutive transformers, λ (inquiring) and α (deciding). In the language of decision theory, λ is also called "experiment," or (in application to medicine), "diagnostic tool." The transformer α is called "rule of action" or "decision rule." The inputs of the "inquiring" box are "events" x and its outputs are "data" y (also called "observations"). The inputs of the "deciding" box are the data y and its outputs are actions a. Thus

$$\lambda(x) = y, \qquad \alpha(y) = a; \quad \text{therefore}$$

$$a = \alpha(\lambda(x)), \text{ or simply } a = \lambda\alpha(x),$$

with the understanding that the last transformation is entered last. Thus, the information system $\lambda\alpha$ has transformed an event x into an action a. The manager must choose from some available (feasible) set of such pairs $\lambda\alpha$, one that is "efficient." Still disregarding for a while the costs associated with each information system, we define the transformer "criterion function" (or, better, gross payoff, or benefit, function) γ, which transforms the input pair (x, a) into the output g, the "gross payoff," or benefit. It depends on the chosen $\lambda\alpha$ thus:

$$g = \gamma(x, a) = \gamma(x, \lambda\alpha(x)).$$

2.2. Events x are, in general, random variables, distributed with ("prior") probabilities π_x; moreover, the transformer λ, "inquiring," and possibly also the transformer α, "deciding," are "noisy," in a sense to be explained presently. As a result, *gross* payoff is also a random variable. By definition, it measures the desirability to the manager of the outcomes of the actions, in the following sense: if costs would not depend on the chosen information system $\lambda\alpha$, he would prefer the system yielding a higher expected payoff to one yielding a lower expected payoff; the word "expected" meaning the average of payoffs weighted by their respective probabilities. These probabilities depend on the "prior" probabilities π_x of events x, and on the conditionsl probabilities characterizing the inquiry λ, and possibly the "deciding" transformer α, as follows.

2.3. Should the inquiring be free of errors, "noiseless," the symbol λ stands for an ordinary function, associating every event x with exactly one observation y. In general, it will not be a one-to-one mapping ("perfect inquiry" is a special, limiting case); rather, it will be a many-to-one mapping: two events, x and x', may yield, for some action a, two distinct payoffs,

$$g(x, a) \neq g(x', a),$$

but the inquiring service may not distinguish between x and x' (it will be "coarser" than a perfect inquiry service):

$$\lambda(x) = \lambda(x').$$

2.4. However, a still more general case is a many-to-many mapping. Then to each $x = x_0$ corresponds not one observation (datum) y, but an array of conditional probabilities $p(y|x_0)$, summing up to 1 over all observations y.

The inquiry λ is then represented by a (Markov) matrix whose rows are such arrays of conditional probabilities of observations, given the events. We shall write $\lambda = [\lambda_{xy}]$ where $\lambda_{xy} = p(y|x)$; λ is called the likelihood matrix. Thus, inquiry λ is, in general, a "stochastic transformation." When we write

$$y = \lambda(x),$$

we shall mean, in general, that the conditional probabilities $p(y|x) = \lambda_{xy}$ are elements of the matrix λ. In the noiseless case, each row of this matrix contains one element 1 (and the rest are therefore zeros); in the "perfect" case, λ is an identity matrix, provided the columns are labelled appropriately.

2.5. We could make analogous statements about the "deciding" service, α. A decision service can be "perfect" (perfectly flexible), or "coarse but noiseless," or "noisy," depending on whether α represents a one-to-one, many-to-one, or many-to-many mapping of data into actions. Intuitively, the reason why a noisy inquiry is chosen is that noiseless (and, even more so, perfect) inquiries are costly, or are not available at all. Similarly a decider (especially if we think not of our ideal manager but of an employee or a machine at his service) may use some non-sophisticated, coarse rules, or may make errors from time to time, and yet be worth hiring if he is sufficiently cheap. But before introducing cost of the inquiring and deciding explicitly, note that the expression $\lambda\alpha$ can be conveniently read as the product of Markov matrices, since, if $a = \lambda\alpha(x)$, then indeed the conditional probability $p(a|x)$ is equal to $\sum_x \lambda_{xy} \alpha_{ya}$, and this is the (x,a)-th element of the product matrix $\lambda\alpha$. The expected gross payoff G, say, can then be written as

$$E(g) = \sum_x \sum_y \sum_a \pi_x \lambda_{xy} \rho_{ya} \gamma(x, a) = G(\lambda, \alpha, \pi, \gamma),$$

where the semicolon separates the entities to be chosen ("controlled") by the manager, from those given to him ("non-controlled"). If costs did not depend on his choice of the information system $\lambda\alpha$, he would maximize the expected gross payoff G over the set of available information systems $\{\lambda\alpha\}$, say. The chosen system (or the set of equally good systems, none worse than any other available one) would depend on the givens, i.e., on the prior probability function π and the gross payoff function γ.

2.6. We can also rewrite the expected gross payoff more explicitly, in terms of the elements of the Markov matrices involved, thus:

$$E(g) = \sum_{xya} \pi_x \lambda_{xy} \alpha_{ya} \gamma(x, a);$$

since the variables x, y, a are "killed" by the triple summation over all their values, G is again seen to depend only on the choice of the information system $\lambda\alpha$ and on the givens π, γ. At the bottom of Figure 1, a simpler expression for the expected gross payoff is given, valid if the deciding service, α, is noiseless (as we may assume for simplicity in what follows).

3. INTRODUCING COST VARIABLE

3.1. Now to the costs. The cost of inquiry depends on the nature of the inquiry (e.g., noisy inquiry is cheaper; a small sample is cheaper than a large one) but also on the particular event that happens to occur. Thus the cost is a function of x,

$$\kappa_\lambda(x), \quad \text{say:}$$

a random variable. (To take sampling again as an example, the cost of a survey of housewives' attitudes to a product will depend on whether the subject was at home on the first visit.) Similarly the cost of a deciding service, e.g., of the decisions to re-order for inventories, will depend on how sophisticated is the re-ordering rule, but also on the random level of the stock at hand, and of the demand predicted by that survey of housewives. Thus the cost of deciding would be, in general a random variable,

$$\kappa_\alpha(y), \quad \text{say.}$$

The expression for the expected cost $E(k)$ is computed at the bottom of Figure 1. Again, it depends, of course, on the information system $\lambda\alpha$ chosen by the manager; and on the prior probability π_x and the cost functions κ_λ, κ_α which are given to him.

3.2. When costs were supposed, temporarily, to be independent of the information system (in 2.2), the manager maximized the expected gross payoff, this being the unique criterion of choice. In this case, so-called utility was identical with gross payoff. Now cost has entered as a second criterion (so that calling γ the "criterion function" is not a good terminology). The utility to the firm, as viewed by the manager (and to be called simply "the manager's utility") is defined as that quantity the expected value of which he tries to maximize by his choice of the information system (which, as we recall, includes deciding as its last component). The utility is now a function of two numerical criteria, gross payoff and cost, increasing in the former and decreasing in the latter. Three cases must be distinguished:

1) Utility is a linear function (a weighted sum) of the two, appropriately scaled criteria, $u(g, k) = wg - k$, with the coefficient (weight, conversion rate) w known; for example, both gross payoff and cost are measured in dollars so that $w = 1$.

2) Utility is a linear function, as above, but w is unknown. In cases 1) and 2) the utility is said to be decomposable (into the component criteria, with respect to each of which it is monotone); if this is not the case we have case 3): $u(g, k)$ cannot be represented as linear in known transforms of g and k.

In case 1) one computes "net expected payoff" as the difference between expected gross payoff G and the expected cost K. Or, a little more generally, one first multiplies one of the criteria by the conversion coefficient w.

In case 2), with w unknown, it is still true that $E(u) = wE(g) - E(k) = G - K$. We say that a choice of information systems that results in G, K __dominates__ another system, which yields G', K', say, if either

$$G \geq G', \quad K < K' \quad \text{or} \quad G > G', \quad K \leq K'.$$

It is clear that in case 2) (and, of course, 1) as well) a system that dominates another system with respect to the expected value of the two criteria will also yield a higher expected utility. One can construct, from the knowledge of the expected values of g and k of all feasible systems, the so-called efficient set consisting of all those feasible systems that are not dominated by some feasible system. All the optimal (but possibly also some non-optimal) systems will be contained in the efficient set. This reduction of the feasible to the efficient set is important in practice. It permits the manager (or his superior, the board, say), to narrow down the choice and to "try out" various values of w: to do some "soul-searching" regarding the conversion rate between benefits and costs, not in the abstract but in the light of concrete possibilities.

3) In case 3), however, an optimal system (i.e., one with maximum expected utility) may have lower criterion expectations than, and thus be dominated by, a non-optimal system. (This mathematical result is due to the fact that expectation is a linear operator.) In this case, our Figures 1, 2, 3, would have to be redrawn. The criterion function would yield directly the utility $u(g, k)$, with the cost k, as one of its inputs, along with x and a.

3.4. In the present three figures, the circles "GROSS PAYOFF" and "COST" have been drawn with "auras" to indicate their dignity as criteria. But in the case of non-decomposable utility (case 3), there would be only one criterion, "utility" itself, to replace "gross payoff," and provided with an aura; the circle "cost" would lose its aura, and have, instead, an output arrow leading from it to the criterion function.

3.5. In most of the practical work, of engineers as well as statisticians and economists, non-decomposable utilities are assumed away, for reasons of simplicity. For this assumption permits to operate with expected values of the individual criterion into which utility is decomposed; and this remains possible, whatever the relevant probability distributions. At some later stage, however, it may become possible to approach more general cases.

4. NO ROLE FOR ENTROPY FORMULAS?

4.1. So far, we have neglected communication. Or, equivalently, we have assumed it to be perfect. That is, we have assumed, in effect, a one-to-one correspondence between the data (observations) put out by inquiring, and the inputs of deciding, which we shall later call, as in Figure 2., "messages decoded." In the context of communication, "entropy formulas" for so-called "information amount" and "capacity" will be introduced. In the context of inquiring and deciding these formulas do not seem to play a role, in spite of numerous writers who have attempted to link the economics of inquiring and deciding with certain results of pure communication economics.

Neither the expected gross payoff nor the expected cost of inquiring and deciding are related to the formulas involving logarithms of the relevant probabilities, as do the entropy formulas.

4.2. In our notation, the entropy formulas depend only on the probabilities π_x and λ_{xy}. In Figure 1 and in Section 2, the expected gross payoff of a system depended, in addition, on the gross payoff function γ and the decision function α. To be sure, the payoff of inquiry alone can be evaluated assuming that the appropriate optimal decision rule is used. We obtain

$$\max_{\alpha} G(\lambda, \alpha; \pi, \gamma) = G_{\pi\gamma}(\lambda), \quad \text{say.}$$

This quantity, sometimes called "value of inquiry," does not depend on α, but still depends on the gross payoff function γ, which will differ from one user of the inquiry service to another. Yet γ does not enter the entropy formulas.

4.3. Suppose the chance of rain a year from now is 50%. Suppose the chance is also 50% that the stock of a corporation in which I hold all my investments will become worthless a year from now. A forecaster whose foresight I absolutely trust offers to tell me whether it will rain or to tell me whether, if I am not careful, I shall lose my fortune. In both cases he will charge \$1,000, arguing that the amount of information he sells is the same in both cases, viz., exactly

$$-(\tfrac{1}{2} \log_2 \tfrac{1}{2} - \tfrac{1}{2} \log_2 \tfrac{1}{2}) = 1 \text{ bit.}$$

Yet, I shall not be indifferent between his two offers. For losing my property is much worse than getting wet: that is, I do take account of the payoff function, when choosing between inquiring services offered to me.

4.4. To illustrate the behavior of cost as well as expected payoff of an inquiry as a function of the matrix, consider the "binary symmetric" case with x and y each taking just two values, labeled 1 and 2, and with

$$\lambda_{11} \equiv p(y=1|x=1) = p(y=2|x=2) \equiv \lambda_{22} \equiv p, \quad \text{say.}$$

Without loss of generality, let p be not less than $\frac{1}{2}$.

It has been shown that the value of inquiry (defined, as we have seen, under assumption of the optimal decision rule), $G_{\pi\gamma}(p)$, say, while depending on π and γ, is non-decreasing in p, regardless of π and γ. (This is plausible intuitively. Remember that 1-p is, in the statisticians' language, the probability of error of either kind.) As to the cost, it is plausible to let it be linear, increasing in the size n of a sample. Interpret events and data as follows:

$x = 1$ or 2 according as the mean of a normal distribution with unit-variance is $+ .1$ or $-.1$;

$y = 1$ or 2 according as the mean of a sample of n is or is not positive.

To achieve a binary symmetric inquiry characterized by p, the sample size n must be equal to $100 \cdot (F^{-1}(p))^2$, where F^{-1} is the inverse of the cumulative normal distribution with mean 0 and unit variance. And the cost of inquiry would be linear in this expression. Again, no relation between an entropy formula and either the value or the cost of inquiry! The value is some non-decreasing function of p depending on the payoff function, while the "amount of information" does not. The cost is a certain increasing convex function of p, again not related to the "amount of information" in any transparent way [nor to the "capacity" of the matrix λ which is $1 + p \log_2 p + (1-p)\log_2 (1-p)$].

5. PURE COMMUNICATION

5.1. As mentioned, the value of becoming informed about future rain and about future loss of my savings is not the same, even if the "amount of information" happened to be the same in both cases. Nor is there any reason to suppose that the cost of obtaining the correct forecast would be the same. What is the same in both cases is neither the value nor the cost of inquiry. Rather, it is the cost of transmitting the message. In both cases, exactly one yes- or -no symbol (one binary digit) needs to be transmitted, corresponding precisely to the number of bits characterizing the probabilities (50-50) of the possible messages. And there is presumably a close relation between the number of bits to be transmitted, and the cost of communication. To transmit 100 binary digits through the same wire one would need 100 times more time-units; or, to use the same time, one would use 100 wires simultaneously, etc.

5.2. The distinction between production and transportation is somewhat analogous. A gallon of whiskey is more costly to produce, and is more enjoyable for the consumer, than a gallon of gasoline. But when it comes to transportation costs a gallon is a gallon. It is quite clear that the originators of the logarithmic formulas of "information theory"--Hartley, Shannon--were fully aware that they were essentially concerned with the cost of communication, not with the cost or value of inquiry. But later writers, impressed by the additive properties of the logarithmic expressions, hailed them as a "measure" of that elusive entity, information, without explaining what the measurement is for. (One recent writer, an expert in the theory of probability, claimed that the measurement permitted to "treat information like money." But there must be some economic reason why we don't measure money by the square feet of the bills' surface!)

5.3. Figure 1, "Inquiring and Deciding" is amplified into Figure 2, "Inquiring, Communicating, Deciding" by inserting, between Inquiring and Deciding, intermediate services, also represented by boxes (i.e., viewed as transformers), and necessary to give account of communication. As a result, the input of the Deciding box is not identical anymore with the output of the Inquiring box. While the latter is (as before) "data," the former is now "messages decoded." Data are transformed into messages decoded through the operations (services) of storing, encoding, transmitting, and decoding, all preceded by storing of the data.

5.4. It is more effective, however, to first present the problems and some results of communication economics (as achieved by the creators of "information theory") by considering the simplified picture given in Figure 3:

"Communication only." It is obtained from Figure 2 by making the following special assumptions:

(1) λ is an identity matrix and $\kappa_\lambda(x)$ is identically zero; no distinction, therefore, between events x and data y.

(2) μ is an identity matrix and $\kappa_\mu(y)$ is identically zero; no distinction therefore between messages to send, m, data y, and (by (1)) events x; that is, "messages to send" (to be denoted by x) enter the criterion function as an input.

(3) α is an identity matrix and $\kappa_\alpha(m')$ is identically zero. Thus action, that is, the other input of the criterion function, is identical with message decoded. Deciding is decoding.

(4) The criterion (or gross payoff or benefit) function has the following form:

$$\gamma(x, a) = \delta_{xa} = \begin{cases} 1 & \text{if } x = a \\ 0 & \text{if } x \neq a. \end{cases}$$

That is, any error is as important as any other. (However, some later writings, following one by Shannon in 1959, drop this assumption and deal with a general "distortion function." I owe this reference to Professor Jacobson of the University of California at San Diego.)

5.5. The encoding function $\epsilon(x) = v$ transforms the (possibly English) message x into a "word" v which is a sequence of symbols (e.g., binary digits) $v_1 v_2 \cdots v_n$, say. Transmitting, symbol by symbol, is done using a "channel" characterized by (a) a Markov matrix τ with as many rows as there are possible input symbols, and as many columns as there are possible output symbols; and (b) the speed of the channel, in symbols per time-unit. The output word put out by the channel is, then, $v' = v_1'v_2'\cdots v_n'$, and the likelihood

$$p(v_1'|v_i)=\tau_{vv'}, \quad \text{(independent of } i\text{)}$$

is an element of the channel matrix τ. We can thus write $v_i' = \tau(v_i)$, with τ a stochastic transformation (as was explained in Section 2 for the analogous case of λ). Finally, the decoding operation d transforms the word v', a sequence of symbols put out by the channel, into a message in the original language. This decoded message, a, together with the original message sent, x, are the inputs of the criterion function which, in most of the literature, is the Kronecker delta, as already mentioned. We have then,

$$a = d\tau e(x),$$

and the gross expected payoff is

$$G = \sum_{\underline{x}}\sum_{\underline{v}}\sum_{\underline{v}'}\sum_{\underline{a}} \pi_{\underline{x}} e_{\underline{xv}} \tau_{\underline{vv}'} d_{\underline{v}'\underline{a}} \delta_{\underline{xa}} = 1 - \textit{Probability of error};$$

(we have underlined the Latin letters to convey that blocks of messages are transmitted)

5.6. On the other hand, there are costs associated with each of the transformers (services) involved. Encoding and decoding costs the more time or effort the larger the length n of the word. And the channel costs the more, the more reliable, in some sense, is its matrix τ, and the greater its speed. No conversion rate is known that would make it possible to express the probability of error, or its complement, in the same units as the cost-determining properties of the code (e, d) and the channel.

5.7. Instead, as Wolfowitz has pointed out, the problem is stated as one of determining the set of non-dominated combinations of G and n (and N, the number of possible words): the efficient set (see case 2) of Section 3). Fundamental is the theorem due to Shannon which states that, provided the "uncertainty at source" is less then the "capacity of channel," a code (e, d)

exists that depresses the probability of error as close to zero as desired; where, in our notation,

uncertainty at source = $-\sum_x \pi_x \log \pi_x$ times the speed of inflow of messages,

capacity of channel = $\max_\rho I(\rho, \tau)$ times the speed of transmission,

where the maximization is over the set of all possible distributions ρ over the alphabet of symbol inputs, and $I(\rho, \tau)$, the "mutual information" of the symbols v_i and v_i' depends only on the probabilities ρ_v, $\tau_{vv'}$, and their logarithms. To achieve a small probability of error with a low-capacity channel, very long code words may be needed. If our problem were not a pure communication problem, and the waiting for the completion of a coded message would imply waiting for a long string of events to happen, the decision would become obsolete. The existence of almost perfect codes would be of no practical interest. In the pure communication situation, however, the messages (not the actual events) do flow in very rapidly. To illustrate: the economics of pure communication is not concerned with following the sequence of events "stock price on Monday, stock price on Tuesday,...," possibly waiting several days to complete an efficiently coded word; rather, it is concerned with transmitting the "stored" record of a long series of such events, or an event rich in dimensions (e.g., the daily Stock Market list of prices). The asymptotic, long-sequence properties of codes and channels, proved in information theory have therefore little relevance, for example, to the economics of sequential decision-making (dynamic programming). Capacity as defined in the theory of communication can be computed for any Markov matrix; but I cannot see that it can be applied usefully outside of the context of coding and transmitting of pre-stored records, except, of course, in fields such as acoustics where the succession of "events" (wave-patterns) is indeed very rapid relative to the needed succession of decisions.

6. INQUIRING, STORING, COMMUNICATING, DECIDING

6.1. We now remove the assumptions (1)-(4) made in the previous section, where the pure communication problem was defined. That is, we shall consider now the sequence of services presented in Figure 2. The gross payoff to the manager depends, not on the messages received compared with the messages sent, but, rather, on the events of the external world, combined with his actions; and his actions do not consist in merely decoding (translating from the language of the channel into ordinary language). Note in particular the transformer "storing" (transforming data into messages, with a time delay necessary to accumulate a "block" of data into an efficiently encodable message). This box did not appear in Figures 1 and 3, where communication was, in effect, separated from the services of inquiry about events, and of decision about actions. Without the storing of data the study of coding and transmitting long sequences of messages, which is the core of the theory of communication, becomes irrelevant to economics.

6.2. The generalization of the expressions for the expected payoff and expected cost that were given in Sections 2 and 3 and at the bottom of Figure 1 is straightforward. The sequence of services, $\lambda\alpha$, becomes now $\lambda\mu\epsilon\tau d\alpha$. (As before, we may consider all these services to be, in general, noisy; if not, the "degeneration" of a Markov matrix into one consisting of 1's and 0's only is easily handled.) We have remarked before that a somewhat "noisy" decider may be cheap. The coding operations (e, d) are often conceived as rigid rules; but we should also think of the complex cases where coding (sometimes called programming in this context) at the present state of technology and of human skills, must be performed as an "art," subject to many trials and errors.

6.3. Although, in Figure 2, a cost function (kappa with an appropriate subscript) is assigned to each service, the accounting practices may or may not have caught up with this task. We have noted, in particular, in Section 5, that the efficient sets that the communication theory strives to construct have dimensions such as "length of code word" (or the expectation of this length), rather than "cost(or expected cost of a code word." In Section 3, we discussed an efficient set of only two dimensions: "expected gross payoff," (or "benefit,") and "expected cost"; the conversion rate between the two being possibly unknown. Perhaps further dimensions must be added pending further research into the monetary cost of coding operations and of prices or rentals of transmitting channels.

6.4. A terminological remark is in order, and should have been made earlier. The manager decides about hiring, among other things, a "deciding service," to be performed by a human or possibly a machine. We have given the example of hiring an employee in charge of deciding about re-ordering for inventories. He must be distinguished from the manager, who makes the "meta-decision" (also called "organizational decision") as to which information services or instruments to use, including the services and instruments for "lower-level" decisions. (It is easy to conceive and philosophize about the infinite recourse of meta-meta-deciders, etc., but we shall not do it here.)

6.5. It is essential to remember that the various services must, in principle, be chosen <u>jointly</u>. The choice of a channel and a code are interdependent, and both are also interdependent with the decision (a "meta-decision" in the sense just defined) as to how detailed or coarse, or how noisy, the inquiry operation should be, and how detailed a message could be typically handled by the deciding employee. The situation is analogous to that of a manufacturer who must decide whether the fuel for his operations should be brought in by rail or by road; this decision must, of course, be made <u>simultaneously</u> with the decision whether to use coal or oil.

6.6. To be sure, it is simpler just to neglect the interdependence between the services constituting an information system. As a first approximation their separability is assumed, and the resulting loss in utility (the "sub-optimization") is accepted. But progress can be expected towards improving the system by taking account of interdependencies between its components. This is quite similar to the progress from a primitive factory design to a modern layout. (Incidentally, the assumption of a "decomposable" utility, linear in the various criteria, such as cost and benefit, is a similar simplification, possibly to be overcome in due course.)

7. DYNAMIC AND STOCHASTIC EXTENSIONS

7.1. So far, the symbols denoting our variables (x, y, m, v, v', m', a, g, k) have not been *dated*, although verbal statements were made as to the time delay involved in storing; and of the various services being more or less costly in terms of time needed for a performance. Account of the processes in time is needed for a proper description of the system and the evaluation of the benefits and costs (which must be "discounted" for time in any calculation of utility)--even in the case when a single decision is to be taken once and for all, as in the case of a simple construction or acquisition project. More usually, the benefits and costs depend on a sequence of decisions, and a sequence of events. A decision to be taken in December will make use of messages about the events of the earlier months of the year, and also take account of the impact of previous decisions.

7.2. Accordingly, one might visualize one sheet such as Figure 3, for each consecutive date, with in-and-output arrows crossing the three-dimensional stack of such sheets. Alternatively, an elaborate network of dated feedback arrows can be used on a sheet. (I believe specialists in information storing and retrieval are working on such problems.)

7.3. In our earlier presentation (as in Section 2.5, for example), the probability distribution of events x was regarded as non-controlled by the manager, as one of the givens in his problem. One approach used in the dynamic programming is to conceive of a sequence of probability distributions, conditional upon the sequence of decisions; the initial, or prior, conditional distribution is followed by a sequence of posterior conditional distributions, revised on the basis of the accumulated sequence of data. Thus only the prior distribution is "given."

7.4. In addition to, or independently, of, this "dynamic extension" of the problem, another extension, or generalization, is often considered, especially by statisticians, and may be called "stochastic." To illustrate, let the action variable have two values: to operate or not to operate on the patient. The events may be "he has cancer," and "not so," with probabilities π_1 and π_2, respectively. However, if he has cancer, the number of years left to him (the "benefit") if action "operate" is taken, is itself a random variable. And the appropriate way to characterize the event "cancer," is to give a probability distribution that will be transformed, if operation is performed, into a certain distribution, and if the operation is not performed, into another distribution of the number of years left to the patient; and similarly, the event "no cancer" is best represented by a probability distribution. Thus the variable x, which influences the "data" x, must be conceived as a "statistical hypothesis," a probability distribution (whether or not conveniently represented by some numerical parameters). Accordingly the benefit and the cost are random variables (whose expectations must be evaluated), not only because the "event" x is subject to a probability distribution π, and because inquiry is, and possibly the other transformers are, noisy, but also because, in general, x itself is a probability distribution (possibly represented by one or more statistical parameters, whose "prior" distribution is given by π).

8. THE MARKET IN INFORMATION SERVICES

8.1. We have assumed the cost functions of the different services to be given. Thus for any given likelihood matrix λ, characterizing an inquiring service, the price $\kappa_\lambda(x)$ is known to the manager. So is the price $\kappa_\tau(v)$ of a given channel matrix τ. (Two channels with the same capacity need not have the same price; for long sequences of messages, they would contribute equally to the expected benefit; and the manager would prefer the cheaper one.) The prices can indeed be considered given to the manager under the regime of competition of numerous firms, facing numerous suppliers, and with no coalitions of suppliers of the information services (thus, no unions of communication workers or computer programmers). If this is not so, then it is not true that the cost functions are given to the manager; he can influence them, depending on his relative bargaining powers. The givens of his problem are, then, these powers (properly defined), and not the prices themselves.

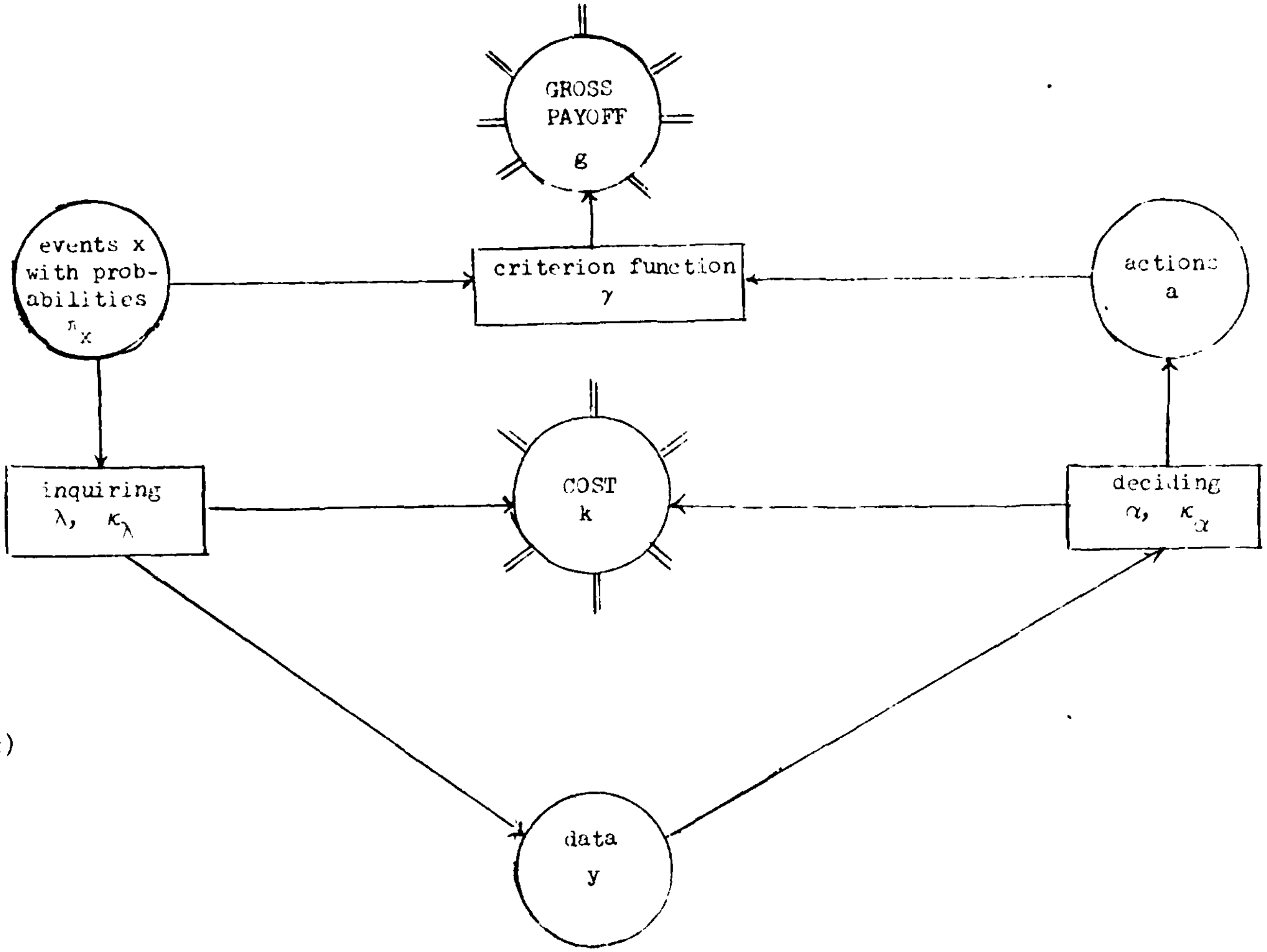
8.2. Whether in a competitive market or not, the price of an information service depends on the way in which the total demand of all managers for a given service depends on its price, and the way in which the total supply of this service depends on this price.

8.3. Our previous discussion explains how the manager <u>should</u> determine his demand for various information services if he has a clear picture of what benefits he wants to achieve. If all managers' ideas of their desired benefits (and also their "prior" ideas about the external world, the distribution π) were known, and if they followed the advice of a management scientist, the total demand for the information services, at any given set of prices, could be evaluated. As it is, one has simply to take some existing state of demand as a fact, a subject of day-to-day market research.

8.4. As to the supply of information services, it depends, of course, on the state of technology (for machines) and of education and training (for men). A comparison between machines and men, and estimation of future trends in their comparative performance, is fascinating and is occupying many minds. As I understand it, compared with the present machines, present man is a very inferior transmission channel and a very poor storer of information. On the other hand, he seems, so far, to be unexcelled in many forms of coding, especially for transmitting to other men (e.g., in efficiencly adjusting the language to the particularities of the receiver), but apparently also to machines (thus, "programming into computers is still an art not a science"-- otherwise it would be all done by machines!) Moreover, current studies in the psychology of language and of information tend to show that, for example, two "inquiring services" that are equivalent in terms of our mathematical definitions (for example, the readings on two instruments with equally fine scales) may be different in an economically relevant sense (vertical scales are read more slowly than horizontal ones). Also, a finer partition of the set of events (e.g., identifying a two-dimensional phenomenon) seems sometimes to require less effort than a coarser partition (e.g., identifying only one dimension), contrary to the guesses that would lead to an easy postulation of an "economic equilibrium."

8.5. It has been estimated that information services as we have defined them constitute 40% or more of the Gross National Product of this country. Hence the public interest in having both the technology and the skills in these fields improved. The purpose of the present paper is merely to contribute to a clearer understanding of the relevant concepts from the point of view of a "manager" (an "organizer," a "meta-decider.")

FIG. 1



$\lambda = [\lambda_{xy}]$,

$\lambda_{xy} = P(y|x)$

Fig. 1. Inquiring and Deciding

$$E(g) = \text{EXPECTED GROSS PAYOFF} = \sum_{xy} \pi_x \lambda_{xy} \gamma(x, \alpha(y)) = G(\lambda, \alpha; \pi, \gamma)$$

$$E(k) = \text{EXPECTED COST} = \sum_{x} \pi_x \kappa_\lambda(x) + \sum_{xy} \pi_x \lambda_{xy} \kappa_\alpha(y) = K(\lambda, \alpha; \pi, \kappa_\lambda, \kappa_\alpha)$$

FIG. 2

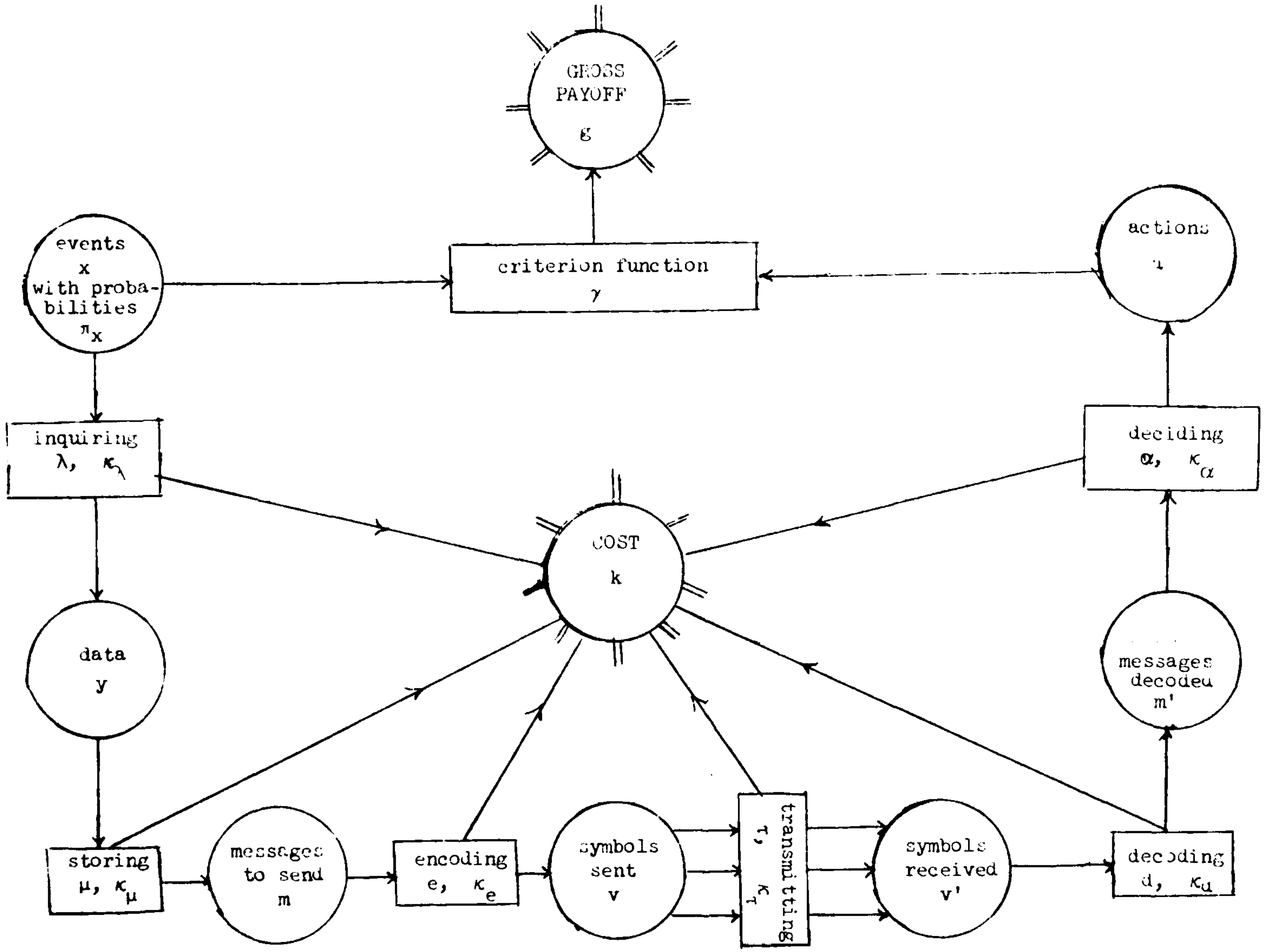
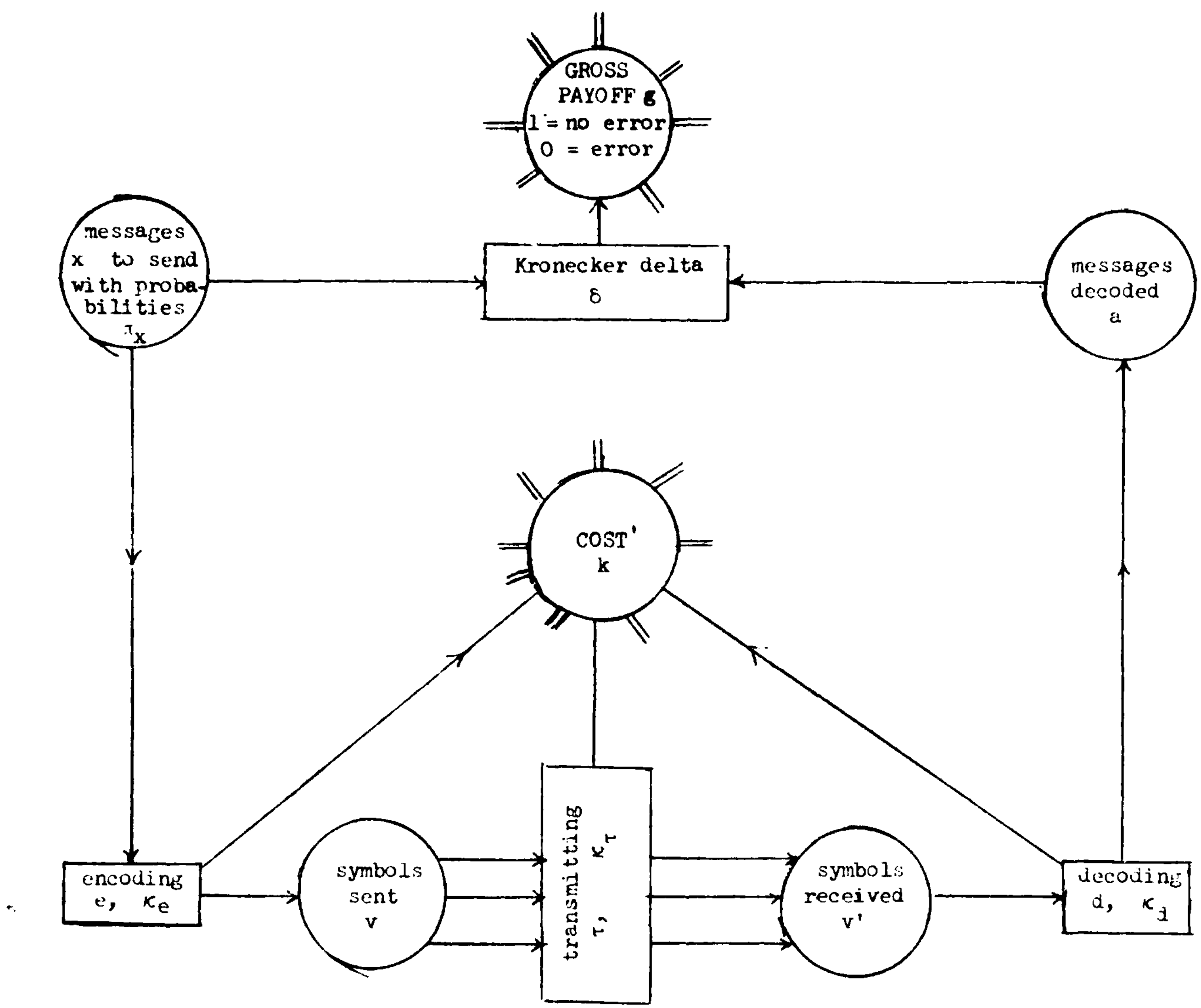


Fig. 2. Inquiring, Communicating, Deciding



Gross payoff $= E(\delta_{xa}) = 1 -$ Probability of error

Fig. 3. Communication only

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author)
Western Management Science Institute

2a REPORT SECURITY CLASSIFICATION
Unclassified

2b GROUP

3 REPORT TITLE
Efficient Choice of Information Services

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

5 AUTHOR(S) (Last name, first name, initial)
Marschak, Jacob

6 REPORT DATE	7a TOTAL NO OF PAGES	7b NO OF REFS
May 1968	23	

8a CONTRACT OR GRANT NO
Nonr 233(75)

b PROJECT NO

c

d

9a ORIGINATOR'S REPORT NUMBER(S)
Working Paper No. 136

9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10 AVAILABILITY/LIMITATION NOTICES
Distribution of this Document is unlimited

Western Management Science Institute
University of California
Los Angeles, California 90024

11 SUPPLEMENTARY NOTES

12 SPONSORING MILITARY ACTIVITY

13 ABSTRACT

An information system is defined as a chain of information services: inquiring--data-storing--encoding--transmitting--decoding--deciding. Each is a transformer represented, in general, by a stochastic matrix and a cost function. The inputs of "inquiring" are the benefit-relevant events (possibly statistical parameters). Actions are outputs of "deciding." Together, actions and events determine the benefits. Other outputs of a service are: (a) inputs into the successive service, and (b) contributions to the cost of acquiring and operating the information system.

The decision theory of economists and statisticians has usually neglected the sub-sequence "data-storing--encoding--transmitting--decoding." Communication engineers, on the other hand, have neglected the inquiring and deciding services and have usually equated benefit with the non-occurence of error in the communication of data. With data pre-stored, long sequences of messages can be communicated without prohibitive delays; and useful asymptotic properties of the "information amount transmitted" and the "channel capacity" follow. These quantities are relevant to the communication cost but neither to the cost nor the benefit of inquiring and deciding.

Suppose the utility to the "manager" (the "organizer," the "meta-decider") is known to be additive in benefit and cost (both appropriately scaled), and his "prior" probability of events is known. Then, and only then, the ("efficient") subset of all feasible information systems for which the pair "expected benefit, expected cost" is not dominated by that of any other system, will contain all

DD FORM 1 JAN 64 1473 0101-807-6800

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
________________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
________________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
________________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

DOCUMENT CONTROL FORM
-continued-

optimal systems. An optimal system can then be determined by a manager compelled to search for, and to apply, his "scaling functions" expressing benefits and costs in the same units.
Correspondingly, pure communication theory has assumed, in effect, utility to be additive in the following criteria (all undesirable, costly, or delay-producing): occurence of communication error; length of code word; size of code; and channel capacity. However, for the efficient choice of the total chain of information services, factors determining the cost of inquiring (e.g., sample size) and of deciding (e.g., computer memory) must also be considered each properly transformed to become an additive component of utility; and an (additive) overall benefit must replace the criterion of "communication error."